Support Vector Machines

Nipun Batra

July 26, 2025

IIT Gandhinagar

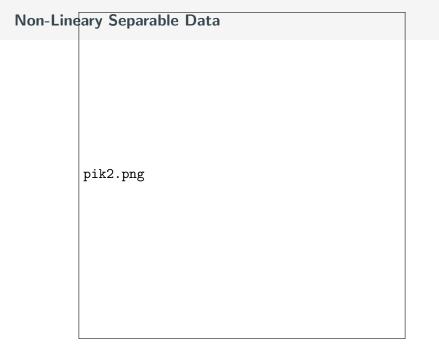
Non-Linearly Separable Data

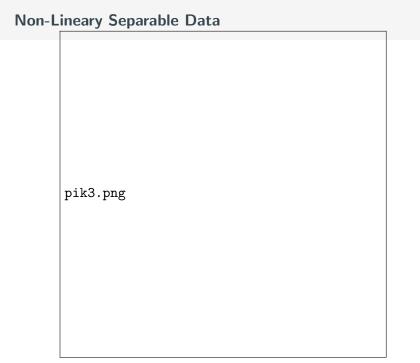
pik1.png

Data not separate in \mathbb{R} Can we still use SVM?

Yes!

How? Project data to a higher dimensional space.





Another Example Transformation pik4.png

Projection/Transformation Function

$$\phi: \mathbb{R}^d o \mathbb{R}^D$$
 where, $d=$ original dimension $D=$ new dimension In our example: $d=1; D=2$

Linear SVM:

Maximize

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \overline{x_i}.\overline{x_j}$$

such that constriants are satisfied.



Transformation (ϕ)



$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \phi(\overline{x_i}).\phi(\overline{x_j})$$

Trivial Example (Again) pik5.png

7 / 100

Steps

1. Compute $\phi(x)$ for each point

$$\phi: \mathbb{R}^d \to \mathbb{R}^D$$

Q. If D >> dBoth steps are expensive!

Steps

1. Compute $\phi(x)$ for each point

$$\phi: \mathbb{R}^d \to \mathbb{R}^D$$

- 2. Compute dot products over \mathbb{R}^D space
- Q. If D >> dBoth steps are expensive!

Kernel Trick

Can we compute $K(\bar{x}_i, \bar{x}_j)$ s.t. $K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i).\phi(\bar{x}_j)$ where, $K(\bar{x}_i, \bar{x}_j) \text{ is some function of dot product in original dimension}$ $\phi(\bar{x}_i).\phi(\bar{x}_j)$ is dot product in high dimensions (after transformation)

Kernel Trick

pik6.png

$$\phi(x)=<\sqrt{2}x,x^2>$$
 $K(x_i,x_j)=(1+x_ix_j)^2-1$ where x_ix_j is dot product in lower dimensions

$$(1+x_ix_j)^2 - 1 = 1 + 2x_ix_j + x_i^2x_j^2 - 1$$

$$= < \sqrt{2}x_i, x_i^2 > . < \sqrt{2}x_j, x_j^2 >$$

$$= \phi(x_i).\phi(x_i)$$



pik7.png

$$\begin{array}{l} \phi(x_1) = <-2\sqrt{2}, 4>; \phi(x_2) = <-\sqrt{2}, 1> \text{ Transformation} \\ \phi(x_1)\phi(x_2) = -2\sqrt{2}\times -\sqrt{2} + 4\times 1 = 8 \text{ Dot product in 2D} \\ \mathsf{K}(x_1, x_2) = \{1+(-2)\times (-1)\}^2 - 1 \text{ Dot product in 1D} \end{array}$$

Kernel Trick

Q) Why did we use dual form? Kernels again!!

Primal form doesn't allow for the kernel trick $K(\bar{x}_1,\bar{x}_2)$ in dual and compute $\phi(x)$ and then dot product in D dimensions

Gram Matrix: (Positive Semi-Definite)

$$K(x_{i}, x_{j}) = (1 + x_{i}x_{j})^{2}$$

$$x_{1} \quad x_{2} \quad x_{3} \quad x_{4} \quad x_{5} \quad x_{6} \quad x_{7}$$

$$x_{1} \quad 24 \quad 8 \quad 0 \quad 0 \quad 8 \quad 24 \quad 48$$

$$x_{2} \quad 8 \quad 1 \quad 0 \quad -1 \quad 0 \quad \dots$$

$$x_{3} \quad 0 \quad \dots \quad \dots \quad \dots$$

$$x_{4} \quad 0$$

$$x_{5} \quad 8$$

$$x_{6} \quad 24$$

48

*X*7

Another Example

.png

Q) What is
$$\phi(x)$$
?

 $\implies \phi(x) = \langle x_1^2, \sqrt{2}x_1x_2, x_2^2 \rangle = x_1^2x_1'^2 + |x_2^2x_2'^2 + 2x_1x_1'x_2x_2'|$

$$K(\bar{x},\bar{x'}) = \phi(\bar{x})\phi(\bar{x'})$$

$$\begin{cases} 2 \\ \Rightarrow (x_1) \end{cases}$$

$$K(\bar{x}, \bar{x'}) = \phi(\bar{x})\phi(\bar{x'})$$

$$K(\bar{x}, \bar{x'}) = \left\{ \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\}^2 = (x_1 x_1' + x_2 x_2')^2$$



14 / 100

Another Example pik9.png

Some Kernels

1. Linear: $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \bar{x}_2$

Some Kernels

- 1. Linear: $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \bar{x}_2$
- 2. Polynomial: $K(\bar{x}_1, \bar{x}_2) = (p + \bar{x}_1\bar{x}_2)^q$

Some Kernels

- 1. Linear: $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \bar{x}_2$
- 2. Polynomial: $K(\bar{x}_1, \bar{x}_2) = (p + \bar{x}_1 \bar{x}_2)^q$
- 3. Gaussian: $K(\bar{x}_1,\bar{x}_2)=e^{-\gamma||\bar{x}_1-\bar{x}_2||^2}$ where $\gamma=\frac{1}{2\sigma^2}$ Also called Radial Basis Function (RBF)

Kernels

Q) For
$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
 what space does kernel $K(\bar{x}, \bar{x'}) = (1 + \bar{x}\bar{x'})^3$ belong to? $\bar{x} \in \mathbb{R}^2$ $\phi(\bar{x}) \in \mathbb{R}^?$
$$K(x,z) = (1 + x_1z_1 + x_2z_2)^3$$

$$= \dots$$

$$= <1, x_1, x_2, x_1^2, x_2^2, x_1^2x_2, x_1x_2^2, x_1^3, x_2^3, x_1x_2 > 10 \text{ dimensional?}$$

Kernels

Q) For $\bar{x} = x$; what space does RBF kernel lie in?

$$K(x,z) = e^{-\gamma ||x-z||^2}$$
$$= e^{-\gamma(x-z)^2}$$

Now:

$$e^{\alpha} = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!}$$

 $\therefore e^{-\gamma(x-z)^2}$ is ∞ dimensional!!

SVM: Parametric or Non-Parametric

Q) Is SVM parametric or non-parametric?

SVM: Parametric or Non-Parametric

Q) Is SVM parametric or non-parametric? Yes and No $\text{Yes} \rightarrow \text{Linear kernel or polynomial kernel (form fixed)}$ No $\rightarrow \text{RBF (form changes with data)}$

$$\begin{split} \hat{y}(x_{test}) &= sign(\bar{w}\bar{x}_{test} + b) \\ &= sign(\sum_{j=1}^{N_{SV}} \alpha_j y_j \bar{x}_j \bar{x}_{test} + b) \\ \hat{y}(X_{test}) &= sign(\sum_{i=1}^{N} \alpha_j y_j K(\bar{x}_j, \bar{x}_{test}) + b) \end{split}$$

 $\alpha_j = 0$ where $j \neq S.V$.

Now $K(\bar{x}_j, \bar{x}_{test})$ for RBF is:

$$e^{-\gamma||\bar{x}_j-\bar{x}_{test}||^2}$$

. Hypothesis is a function of "all" train points

Closer \bar{x} is to \bar{x}_N ; more is it influencing $\hat{y}(\bar{x})$ - hypothesis function

pik10.png

• Now if we add a point to the dataset

- Now if we add a point to the dataset
- Functional form can adapt (similar to KNN)

- Now if we add a point to the dataset
- Functional form can adapt (similar to KNN)
- .: SVM with RBF kernel is non-parametric

•
$$\hat{y}(x) = sign(\sum \alpha_i y_i e^{-||x-x_i||^2} + b)$$

•
$$\hat{y}(x) = sign(\sum \alpha_i y_i e^{-||x-x_i||^2} + b)$$

 \bullet $-||x-x_i||^2$ corresponds to radial term

•
$$\hat{y}(x) = sign(\sum \alpha_i y_i e^{-||x-x_i||^2} + b)$$

- $-||x-x_i||^2$ corresponds to radial term
- $\sum \alpha_i y_i$ is the activation component

•
$$\hat{y}(x) = sign(\sum \alpha_i y_i e^{-||x-x_i||^2} + b)$$

- $-||x-x_i||^2$ corresponds to radial term
- $\sum \alpha_i y_i$ is the activation component
- $e^{-||x-x_i||^2}$ is the basis component

RBF: Effect of γ

 γ : How far is the influence of a single training sample

pik11.png